Face Image Recognition Based on Convolutional Neural Network

Guangxin Lou, Hongzhen Shi*

Yunnan Minzu University, Kunming 650500, China
* The corresponding author, email: yingbinggan@126.com

Abstract: With the continuous progress of The Times and the development of technology, the rise of network social media has also brought the "explosive" growth of image data. As one of the main ways of People's Daily communication, image is widely used as a carrier of communication because of its rich content, intuitive and other advantages. Image recognition based on convolution neural network is the first application in the field of image recognition. A series of algorithm operations such as image eigenvalue extraction, recognition and convolution are used to identify and analyze different images. The rapid development of artificial intelligence makes machine learning more and more important in its research field. Use algorithms to learn each piece of data and predict the outcome. This has become an important key to open the door of artificial intelligence. In machine vision, image recognition is the foundation, but how to associate the low-level information in the image with the high-level image semantics becomes the key problem of image recognition. Predecessors have provided many model algorithms, which have laid a solid foundation for the development of artificial intelligence and image recognition. The multi-level information fusion model based on the VGG16 model is an improvement on the fully connected neural network. Different from full connection network, convolutional neural network does not use full connection method in each layer of neurons of neural network, but USES some nodes for connection. Although this method reduces the computation time, due to the fact that the convolutional neural network model will lose some useful feature information in the process of propagation and calculation, this paper improves the model to be a multi-level information fusion of the convolution calculation method, and further recovers the discarded feature information, so as to improve the recognition rate of the image. VGG divides the network into five groups (mimicking the five layers of AlexNet), yet it USES 3*3 filters and combines them as a convolution sequence. Network deeper DCNN, channel number is bigger. The recognition rate of the model was verified by 0RL Face Database, BioID Face Database and CASIA Face Image Database.

Keywords: convolutional neural network; face image recognition; machine learning; artificial intelligence;multilayer information fusion

I. OVERVIEW

Nowadays, with the development of modernization, people live in a fast-paced way,

Received: Jun. 28, 2019
Revised: Nov. 3, 2019
Editor: Zan Li

The multi-level information fusion model based on the VGG16 model is an improvement on the fully connected neural network.

usually replaced by a picture in the process of communication. From the perspective of users, images can better express and convey a person's emotional information and serve as a carrier for rapid expression, because they are not restricted by local language and other factors. It is easy for us to edit text, but the content contained in images and the amount of information transmitted are so large that how users can find the key information they need and apply it conveniently and efficiently has become a key research problem in the field of image processing and data analysis[1].

In recent years, image recognition has become a hot topic in the field of machine vision and image research due to its wide use of scene. In animal husbandry, we can judge by monitoring the growth situation of livestock, body shape and shape, feed and feed livestock regularly and quantitatively, and monitor their growth and development; In industry, each key point can be monitored by image recognition. In medicine, patients can be screened for various parts of the body through high-definition images. Space, according to the satellite real-time transmission of image data, research; In addition, images also play a vital role in our daily activities, such as face recognition in the security check of high-speed trains and airlines to determine the identity information of passengers; When parking on weekdays, self-access to the parking lot, using license plate recognition technology; when applying for an id card, each citizen will input his or her fingerprint information for future security verification. Google's virtual reality and driverless technology also uses incoming image recognition. In a word, image is indispensable in our daily life. It is of great significance to use traditional methods [2-3] or deep learning [4-5] to quickly and effectively identify images and extract relevant feature information, such as color feature [6]. Image recognition based on convolutional neural network is the first application in the field of image recognition. A series of algorithm operations such as extraction, recognition and convolution operation of image eigenvalues are used to identify and analyze different images, and its recognition effect is generally recognized and affirmed by scholars.

In machine vision, image recognition is a key problem. The accuracy of image recognition mainly depends on the accuracy of the positioning of key parts of the object and the learning degree of relevant features [7 -8]. First, image processing is image preprocessing [9]. After pretreatment, image features are analyzed and targets are identified. Currently, feature extraction algorithms are mainly SIFT [10], LBP [11] and Haar [12]. Through a series of steps, the content in the image can be accurately identified. There are many ways to get the image, but it is essential to use the photography tool to store the captured image data in the database for easy use and management. Image processing is to cut the image we need in the gallery to determine its uniform format and size. For images with poor image quality, the image can be denoised, enhanced and a series of preprocessing, which is convenient for future use. Getting basic functional information from the external form of the image is a step called feature extraction. Obtaining complete feature information is the key to truly reflect the needs of users, thus serving the industry, so this step is the basis of the image identification process. Global features include color, shape, texture, etc. [13]. Good processing of global features will produce better visual effects for image recognition. Local features include prominent structures, such as corners, edges, and patches. Global features and local features make up the big definition of features. The two meanings are the criteria for constructing the classifier, and also the computer to describe the form of the picture. First, extract the features of the image to be recognized [14], and divide these features into different groups through the classifier, so as to finally achieve the purpose of image recognition.

The development of AI and machine learning also provides superior resources for image recognition. Machine learning [15] is to let machines have brains like human beings, so

that machines can independently learn and analyze problems. Human knowledge acquisition is a long and repeated process. In the process of human evolution, people continue to learn and understand knowledge so as to understand what they need to acquire. Through constant practice and failure, they get relevant skills from it and analyze and summarize experience in the process of continuous practice. Finally get what you want. Machine learning is similar to human brain learning, enabling machines to learn on their own is no longer a scene from science fiction movies. Humans capture all the information around them through their own eyes, and transmit retinal imaging to the brain to obtain relevant image features. They stimulate cells in the brain to form their own thoughts, and apply a series of characteristics such as behaviors and actions to things they have learned. Machine vision [16] is to obtain relevant images through equipment. After feature extraction, the machine can eventually identify the relevant object through learning [17-18] and present what it has learned through a computer or audio device. Classifier training is the key to machine learning. The classifier model is obtained by training and adjusting the classifier continuously, which can identify the objects with small error range. Machine learning includes unsupervised learning, supervised learning and semi-supervised learning [19].

Supervised learning: manual intervention is required in the learning process. In the database, the image data contains labels, and the classifier is further adjusted by analyzing the results. Supervised learning is the earliest machine learning method created by humans.

Semi-supervised learning: after the classifier is generated, part of the data in the database is tagged and the other part is not tagged, and the tagged image data is adjusted to achieve the purpose of correction.

Unsupervised learning: all image data in the database does not contain labels and does not involve manual intervention. The computer completely completes the simulation human brain process independently, finally achieves the study goal.

As an important part of machine learning in the field of artificial intelligence [20], deep learning, which is gradually developed to imitate human visual system, is the most prominent field in this field. Eyes are the window of people's mind. The color, size, shape and other visual information of everything can stimulate people's visual nerve through imaging, activate a series of neurons, and finally transmit signals to the brain. This process is simple to say, but it takes different parts of the nervous system to transfer and abstract the concept of objects in the brain. Human learning requires constant repetition. The same is true for the transmission of neural signals. The lowest level of features are transmitted to the next level, and each layer is abstracted and transmitted. This process is carried out continuously, and only through continuous training can the high-level meaning signals be finally converted. Deep learning is the earliest application in the field of image processing, and the process is similar to the process of object recognition by human beings. By constantly abstracting semantics and iterating to identify objects with the pixel information at the lowest level in the image, the higher the level, the higher the accuracy of identifying objects will be.

With the rapid development of interconnection, a huge amount of image data also follows [21]. Machine learning has the advantages of high recognition rate and strong expansibility. Continuously create and improve the learning model, seek for valuable image information, and further bring more innovation and opportunities for image recognition, so as to better serve human beings and promote the continuous progress and development of society.

Nowadays, Image processing has become a popular topic. Face recognition technology is widely used in image processing [22]. How to quickly and efficiently recognize image information and process relevant features has become a hot focus. Applying machine learning to image recognition can greatly speed up the recognition efficiency and improve the accuracy. In machine learning, the most tradi-

tional model is convolutional neural network model [23]. In the ILSVRC-2012 competition, the convolution neural network algorithm [24] has achieved excellent results. The model has a series of advantages, but some useful image information will be discarded in the forward propagation process of convolutional neural network, which cannot be well utilized as image features. Therefore, in this paper, the discarded image features are recycled and utilized in the design process, and the multi-layer information fusion method is used to improve the traditional convolutional neural network model, and the face image recognition is studied by using this model.

II. Convolutional Neural Network

Convolutional neural network has been proposed in the early stage, and the rapid development of artificial intelligence in recent years has brought this model back to the perspective of scholars. It plays an important role in image technology, biomedical technology, industrial production and other fields. The earliest concept of neural network is extended by scientists in the human nervous system [25], which imitates the human nervous system and puts forward the concept of neural network. Convolutional neural network is further improved from the concept of neural network [26]. The emergence of this model brings good news to machine vision.

III. Lenet-5 Model

lenet-5 model is made up of Convolutional Layer, Pooling layer and Fully Connected Layer. The network diagram is shown in figure 1.

IV. Convolution Neural Network Based on VGG16

VGG divides the network into five groups (mimicking the five layers of AlexNet), yet it USES 3*3 filters and combines them as a convolution sequence. Network deeper DCNN, channel number is bigger [27]. The network structure is shown in figure 2.

In the vgg16 model [28], the convolution kernel of the filter is 3*3*3, and then 64 convolution kernels ((3*3*3) *64) are used to convolve with the input image, and the corresponding output of the first convolution layer is 224*224*64.

The convolution of the first layer is output to the convolution of the second layer. The second layer has 64 convolutions respectively with the previous layer, and the output of the second layer is 224*224*64.

There is a pooling layer between the output of the third convolution layer and the

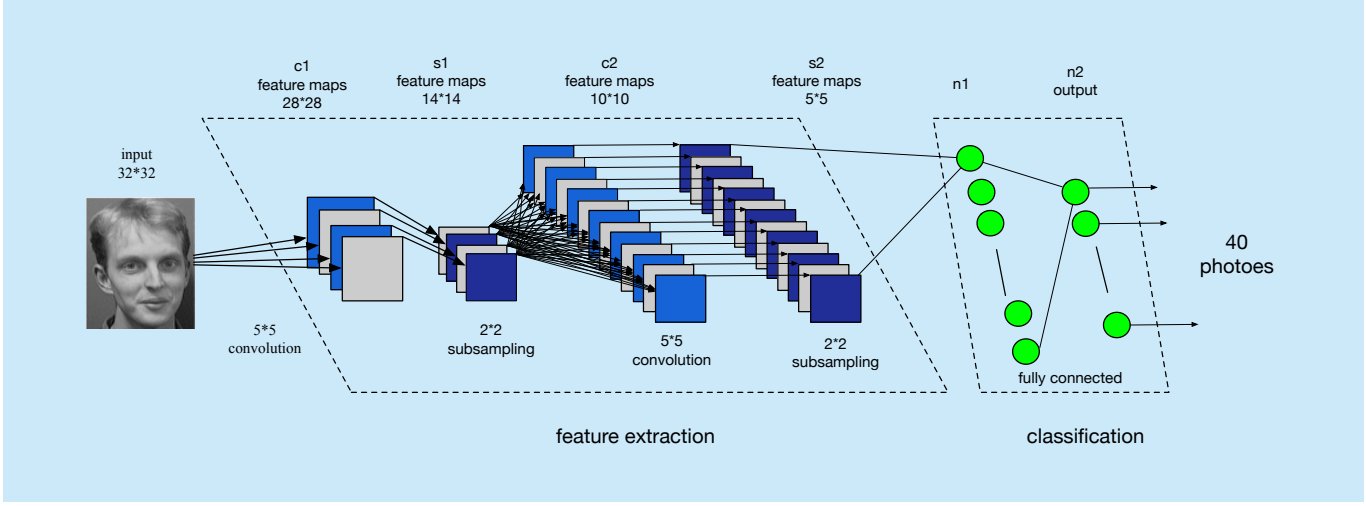

Fig. 1. *Lenet-5 model.*

output of the fourth convolution layer. Since the pooling layer will reduce the length and width of the input of the previous layer but will not increase the number of channels, the output of the first red pooling layer above is 112*112*64.

In vgg16 model, convolution is used instead of full connection, mainly because this method can improve the efficiency of calculation, reduce the consumption of memory, and do not limit the size of the input image. In the traditional CNN convolutional network, the number of parameters in the first few layers of convolutional layer occupy a small proportion and the calculation amount occupies a large proportion. The latter full connection layer is the opposite. Therefore, we focus on the convolution layer when carrying out computational acceleration optimization; In the process of parameter optimization and weight clipping, the research focus is on the full connection layer.

V. Face Recognition Based on Convolution Neural Network

5.1 ORL face database

Created by AT&T Laboratory of Cambridge University, it contains 400 facial images of 40 persons with different facial expressions. The differences are as follows [29]:

(1) The same person shows different postures

(2) The same person's facial expression, with different ornaments

(3) The same person in different light and different environment, which results in different gray values of the image. The images of 40 testers are listed in figure 3.

5.2 BioID Face Database

This dataset contains 1,521 grayscale images with a resolution of 384x286 pixels. Each image was taken from a face at a face Angle of 23 different subjects. It was recorded under different illumination and complex background. The eye position is manually set.

5.3 CASIA Face Image Database

CASIA Face Image Database Version 5.0 contains 2,500 color facial images of 500subjects.

5.4 Algorithm Improvement

The traditional convolutional neural network

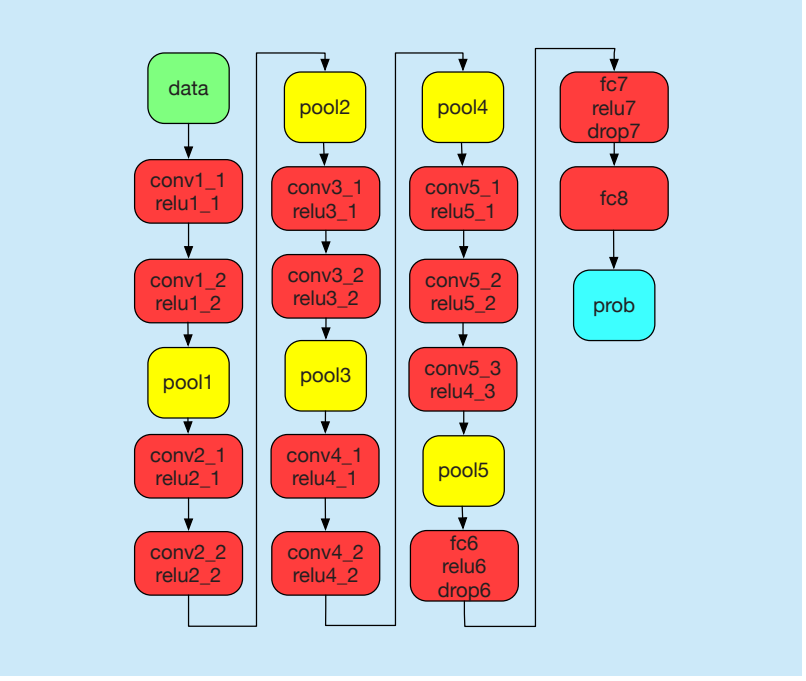

Fig. 2. *VGG16 network structure.*

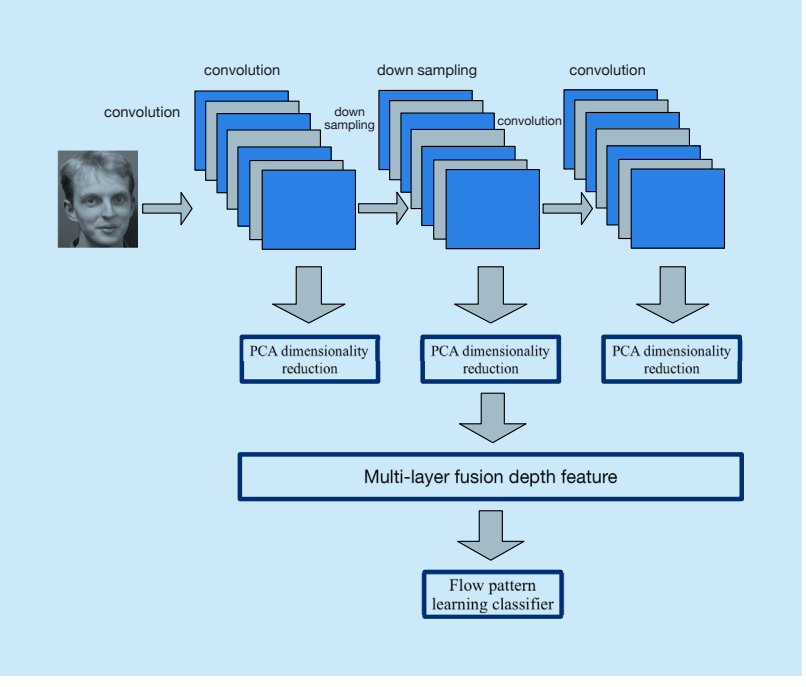

Fig. 3. *Improved algorithm.*

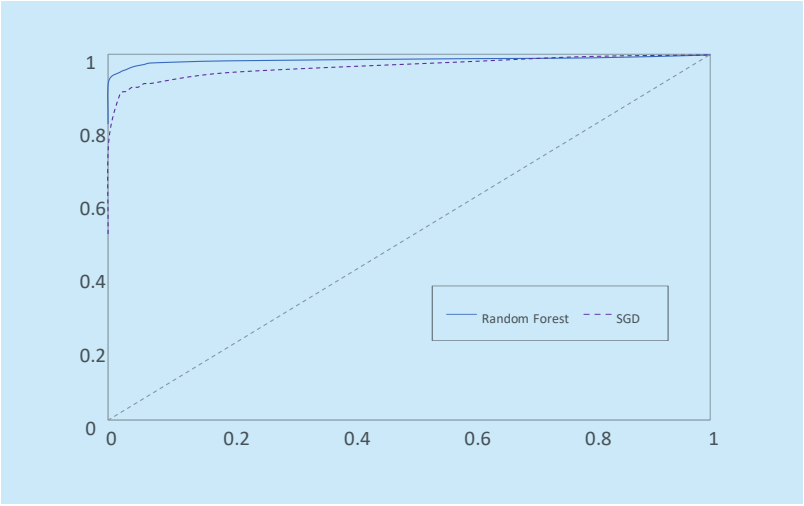

Fig. 4. *The ROC curve of ORL face database.*

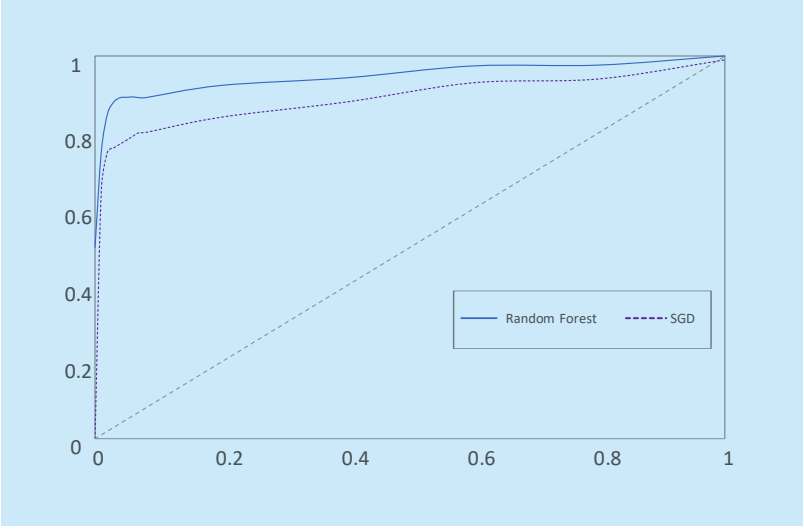

Fig. 5. *The ROC curve of BioID face database.*

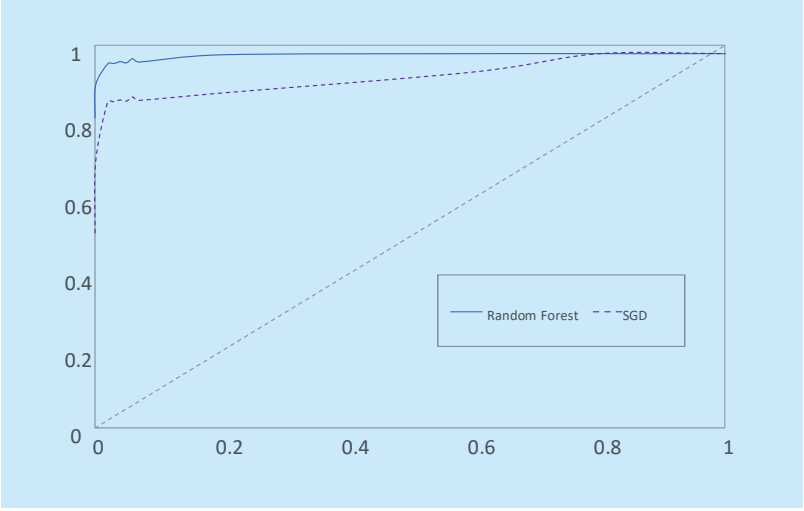

Fig. 6. *The ROC curve of CASIA face image database.*

algorithm is similar to image mapping, which filters image features layer by layer and extracts features from the last layer as classification. After each layer of filtering, some useful data is left because the information is not fully utilized. These eigenvalues can be extracted and calculated to improve the recognition rate of the image. The network structure of information fusion is to conduct feature fusion between the specific convolution layer before the pooling layer of VGG16 and the original image. The improved principle is shown below.

Output feature X_n^m. in the first convolution layer of vgg16 n the convolution layer, the feature image of m-1 layer is convolved with each one and carried out convolution operation. After the activation function operation, the feature graph of m layer is obtained. Convolution operation is as follows:

$$X_n^m = f\left(\sum_{i\in M_n} X_i^{m-1} \times K_{in}^m + b_n^m\right), \tag{1}$$

m represents the current convolution layer, K_{in}^m represents the convolution kernel connecting the i-th feature graph at the m-1st layer with the n-th feature graph at the m-1st layer. b_n^m represents that the bias f of the n-th feature graph at the m layer is the activation function. X_i^{m-1} represents the i-th image feature on the m-1st layer. The feature graph can be deconvolved to obtain the same feature as the original graph. Where, the deconvolution formula is as follows:

$$\Sigma_{n=1}^{k1} X_n^m \otimes f_{j,c} = y_c. \tag{2}$$

The c dimension characteristic graph is obtained by linear summation of ki characteristic graphs. The feature image after deconvolution is fused with the input image.

The activation value output by vgg16 structure and the activation value in multi-layer information fusion network were added and calculated by the center loss function and SoftMax loss function [30]. The formula is as follows:

$$L = L_s + \lambda L_c$$

$$= -\sum_{i=1}^{m} \log \frac{e^{W_{yi}^T \lambda_i + b_{yi}}}{\sum_{j=1}^{n} e^{W_j^T \lambda_i = b_i}} + \frac{\lambda}{2} \sum_{i=1}^{m} \left\| x_i - c_{yi} \right\|_2^2. \tag{3}$$

5.4 Experimental analysis

5.4.1 ROC curve

According to the curve of ROC curve, FPR and TPR under different thresholds are shown. The improved convolutional neural network has excellent performance in face recognition. The value of AUC can be perfectly approximated to 1.

5.3.2 Results and comparisons

It can be seen from the comparative experiment that the structure of convolutional neural network has a high advantage. The comparison results are shown in table 1.

As can be seen from the table, compared with the five algorithms, the image recognition rate of the three data sets in this paper is outstanding, about 5.27% higher than lCA algorithm [31], about 1.5% higher than Eigenface, about 0.7% higher than 2DPCA, and about 0.51% higher than Fisherface. The improved image recognition rate algorithm is 0.27% higher than that of convolution neural network. The improved model has good recognition effect on ORL face image. Although the recognition rate of BioID data set and CASIA data set is not as good as that of ORL data set, compared with other algorithms, VGG16 multi-layer fusion convolution model can further improve the recognition rate of images.

Table I. *Comparison of experimental results.*

Recognition Algorithm	Recognition Accuracy of ORL /%	Recognition Accuracy of BioID /%	Recognition Accuracy of CASIA/%
Eigenface	97.50	97.41	97.06
Fisherface	98.50	96.57	96.44
ICA	93.75	93.41	93.62
2DPCA	98.30	96.31	97.21
Convolution Neural Network	98.75	96.53	97.89
Convolution Neural Network of This Paper	99.02	97.62	98.65

VI. Conclusion

Convolution neural network, as a common model structure in machine learning, has achieved excellent results in image enhancement, image fusion, image processing, image recognition and so on, and is favored by experts and scholars. The discarded image information is recycled and utilized on the original convolutional neural network model, and the improved model can further improve the performance of machine self-learning and classification. Especially in the image recognition, after a series of processes such as the calculation, pooling and abstraction of convolutional neural network, the machine can accurately identify the image feature information. In the future, there are many advances and improvements in the development of machine vision.

References

[1] D, Han, research on image retouging algorithm based on sparse representation. Diss. Xi 'an Danzi university, 2013.

[2] G. Guo and M.J. Jones. "iris extraction based on intensity gradient and texture difference." *2008 IEEE Workshop on Applications of Computer Vision*, IEEE, 2008, pp. 1-6.

[3] F. Perronnin et al. "improving the fisher kernel for large-scale image classification." *European conference on computer vision*, Springer, 2010, pp. 143-156.

[4] E. Gavves et al. "local alignments for fine-grained categorization." *International Journal of Computer Vision*, vol. 111, no. 2, 2015, pp. 191-212.

[5] C. Pang et al. "part-aware segmentation for fine-grained categorization." *Pacific rim conference on multimedia*, Springer, 2015, pp. 538-548.

[6] S. Bloehdorn and A. Hotho. "boosting for text classification with semantic features." *International workshop on knowledge discovery on the web*, Springer, 2004, pp. 149-166.

[7] Y. Gao et al. "compact bilinear pooling." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 317-326.

[8] G. Zheng et al. "fine-grained image recognition via weakly supervised click data guided bilinear cnn model." *2017 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2017, pp. 661-666.

[9] R. Chen, et al. "research on image matching of machine vision." *Modern Electronics Technique*, 2011.34(18):46-49+52.

[10] D.G. Lowe. "object recognition from local scale-invariant features." *iccv*, vol. 99, 1999, pp.

1150-1157.
[11] t. Ojala et al. "gray scale and rotation invariant texture classification with local binary patterns." *European Conference on Computer Vision*, Springer, 2000, pp. 404-420.
[12] C.P. Papageorgiou et al. "a general framework for object detection." *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, IEEE, 1998, pp. 555-562.
[13] A. Krizhevsky et al. "imagenet classification with deep convolutional neural networks." *Advances in Neural Information Processing Systems*, 2012, pp. 1097-1105.
[14] Y. LeCun et al. "deep learning." *Nature*, vol. 521, no. 7553, 2015, pp. 436-444.
[15] C.B. Zhu et al. "research on image classification algorithm based on semi-supervised deep belief network." *Computer Science*, 2016.
[16] C. Zhu, et al. "development and application of computer vision technology." *Heilongjiang science and technology information*, vol. 29, 2013, pp. 174.
[17] W. Wang, et al. "image aesthetic classification based on parallel deep convolutional neural network." *Chinese journal of automation*, vol. 42, no. 6, 2016, pp. 904-914.
[18] Q. Li et al. "medical image classification with convolutional neural network." *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, IEEE, 2014, pp. 844-848.
[19] T. Randen and J.H. Husøy. "filtering for texture classification: a comparative study." *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 4, 1999, pp. 291-310.
[20] Y Chen. improvement of deep convolutional neural network algorithm [J]. Electronic world,2018(07):57-58.
[21] X. Chang. research on image classification application based on convolution neural network nozzle. Diss. North University of China, 2017.
[22] A. Krizhevsky et al. "imagenet classification with deep convolutional neural networks." *Advances in Neural Information Processing Systems*, 2012, pp. 1097-1105.
[23] D.H. Hubel and T.N. Wiesel. "receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *The Journal of physiology*, vol. 160, no. 1, 1962, pp. 106-154.
[24] Y. Le Cun, L. et al. "gradient-based learning applied to document recognition." *Proceedings of the IEEE*, vol. 86, no. 11, 1998, pp. 2278-2324.
[25] Y. Bengio et al. "representation learning: a review and new perspectives." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, 2013, pp. 1798-1828.
[26] Y. Li, research on key technologies of computer vision based on convolution neural networks nozzle. Diss. University of Electronic Science and Technology of China, 2017.
[27] K. Simonyan and A. Zisserman. "two-stream convolutional networks for action recognition in videos." *Advances in Neural Information Processing Systems*, 2014, pp. 568-576.
[28] Y, Cai, et al. classification of trash types in cotton based on deep learning. MS thesis. Wuhan University Of Science And Technology, 2019.
[29] S. Zhang. research on face recognition based on large-scale database. Diss. Beijing Jiaotong University, 2009.
[30] Y. Wen et al. "a discriminative feature learning approach for deep face recognition." *European conference on computer vision*, Springer, 2016, pp. 499-515.
[31] W. Gao, et al. "application research of image recognition based on convolutional neural network algorithm." *Journal of xi 'an university of arts and sciences (natural science edition)*, vol. 22, no. 1, 2019, pp. 10-14.

Biographies

Guangxin Lou, is undergraduate student in Yunnan Minzu University . His research interests include image classification and image recognition. Email:lou123guangxin@163.com

Hongzhen Shi, is a lecturer in the department of information engineering at the Yunnan Minzu university .Her research interests include image classification, Image inpainting. Email: yingbinggan@126.com